Explaining cross-cultural differences

Poortinga, Y.H.; van de Vijver, F.J.R.

Published in:
Journal of Cross-Cultural Psychology

Publication date:
1987

Citation for published version (APA):
Poortinga, Y. H., & van de Vijver, F. J. R. (1987). Explaining cross-cultural differences: Bias analysis and beyond. *Journal of Cross-Cultural Psychology*, *18*(3), 259-282.

Contrary to the tradition in cross-cultural psychology in which bias is identified with the stimulus by culture interaction in an analysis of variance design, it is argued (and illustrated) here that any main effect or interaction term in such a design can reflect bias. It is suggested that difficulties in the interpretation of observed cross-cultural differences in data can be avoided only if the researcher succeeds in introducing external variables in terms of which the variance in the factor culture can be reduced to zero. This implies, somewhat paradoxically, that in a satisfactory cross-cultural study there is no variance left to be explained in terms of culture.

EXPLAINING CROSS-CULTURAL DIFFERENCES
Bias Analysis and Beyond

YPE H. POORTINGA
FONS J. R. VAN DE VIJVER
Tilburg University

The interpretation of cross-cultural differences in performance on psychological measurements involves a serious dilemma. Are the results a valid indication of differences in psychological functioning between groups of humans, or should they be explained in terms of bias or incomparability of the data? The most obvious example to illustrate the dilemma is the controversy on the measurement of intergroup differences in intelligence test scores. The egalitarians maintain that unequal mean scores on tests reflect differences in concomitant variables such as test-taking attitudes and the opportunity to acquire information needed to perform well on the test. In other words, they maintain that the tests measure nonidentical psychological constructs in culturally different populations. On the basis of the same results, nonegalitarians hold that inferences about

AUTHORS' NOTE: This article was written while the first author was a Fellow at the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS), Wassenaar, The Netherlands.

JOURNAL OF CROSS-CULTURAL PSYCHOLOGY, Vol. 18 No. 3, September 1987, 259-282

cross-cultural differences in intelligence are justified, for example, by arguing that the performance differences reflect genetic intergroup differences. This implies that the tests are supposed to measure identical aspects of behavior across cultures.

In the psychological literature, the debate between these conflicting views has primarily centered around intellectual abilities and their measurement, although the problem of incomparability is equally pertinent to other areas of assessment, such as the measurement of personality traits and social interaction processes. For example, in their cross-cultural studies of affective meaning, Osgood and his associates found in the United States an unusually high positive evaluation for being aggressive. Rather than concluding that people in this country highly value aggressiveness in the usual sense of being intentionally injurious to others, Osgood (e.g., 1977, p. 231) opted for a different interpretation, pointing out that in the United States being aggressive also means being competitive in school or sports. If Osgood is correct, a comparison of ratings on "aggressiveness" in terms of the intended injury concept between the United States and some other cultural group would lead to wrong conclusions.

During the last decades, important developments have taken place in the debate. More than before, the issue of score incomparability and cultural bias has become the subject of empirical scrutiny. Developments in item bias analysis are of particular relevance here. Quite some effort has been put into developing and refining psychometric procedures to detect incomparability of items (e.g., Berk, 1982; Lord, 1980; Mellenbergh, 1982; Van der Flier, 1982; Van der Flier, Mellenbergh, Adèr, & Wijn, 1984). An attractive feature of these procedures is that they are independent of the prior beliefs of the investigator, thereby introducing notions of falsifiability and replicability.

Apart from these advantages, bias detection procedures also have important shortcomings, which have received insufficient attention in our opinion. The interpretation dilemma cannot be resolved completely by bias analysis. Beyond the identification of bias a further step is needed, namely, the *explanation* of cross-cultural differences. To this purpose, explanatory variables should be included in the design of a study. We shall present a hierarchical regression model as a framework that allows for a coherent analysis of the major issues involved in the interpretation of cross-cultural differences, provided the explanatory variables meet certain constraints.

THE IDENTIFICATION OF BIAS

The design of many cross-cultural studies can be represented by analysis of variance models (see Van de Vijver & Poortinga, 1982). In the most simple case, a set of stimuli, for example, a test, has been administered to persons belonging to different cultural groups. In such a case the analysis of variance (ANOVA) model is given by:

$$X_{sp(c)} = \mu + C_c + S_s + P_{p,}\ PC_{pc} + SC_{sc} + SP_{sp,}\ SPC_{spc,}\ E_{spc} \qquad [1]$$

where:

- μ is the overall mean;
- C_c ($c = 1, \ldots, n_c$) is the main effect for culture;
- S_s ($s = 1, \ldots, n_s$) is the main effect for stimuli;
- $P_{p,}\ PC_{pc}$ ($p = 1, \ldots, n_p$) is the confounded effect of the main effect for persons and the person by culture interaction;
- SC_{sc} is the interaction between stimulus and culture;
- $SP_{sp,}\ SPC_{spc,}\ E_{spc}$ is the confounding of the stimulus by person interaction, the stimulus by person by culture interaction, and the error term (E).

Frequently, the stimulus by culture interaction is interpreted as bias (e.g., Cleary & Hilton, 1968; Poortinga, 1971). If this interaction is small and statistically not significant, the main effect for culture tends to be conceived of as an index of a valid cross-cultural difference, that is, as indicating a cross-cultural difference in the domain of behavior or the trait to which the scores are generalized. This line of reasoning has a high intuitive appeal. Consistent cross-cultural differences should be reflected by most if not all stimuli in an instrument, whereas cultural differences reflected by only a few are more likely a consequence of exceptional stimuli that tap somewhat different constructs in various cultures.

It should be noted that the ANOVA model is used here as an example. It may not always be the most adequate framework for the identification of bias. Lord (1980) has shown that it can have certain disadvantages. For instance, the SC-interaction can reflect measurement artifacts when floor or ceiling effects are present in the scores. So-called latent trait models (Lord, 1980) or χ^2 approaches (Marascuilo & Slaughter, 1981) have fewer shortcomings. However, the comments we shall make apply also to these psychometric models used for bias analysis.

A CRITIQUE OF BIAS ANALYSIS

The traditional procedures for bias analysis are based on the assumption that the sources of inequivalence affect the data only at item level. Bias is supposedly unambiguously traced in a data set by searching for items that perform differentially across cultures. Consequently, it is assumed that there are no bias factors that affect all items to a similar extent and, by implication, the main effect for culture in an ANOVA design. This assumption provides the theoretical background for the interpretation of the C-component as reflecting true differences and the SC-component as bias.

This interpretation is often debatable, both on theoretical and psychometric grounds. To start with the former, it is unrealistic to assume that cross-cultural psychologists will always consider the SC-interaction as bias. A researcher may be interested in the SC-interaction as an index of cultural differences, rather than in the main effect for C.

Take, for example, the cross-cultural work on choice reaction times (e.g., Jensen, 1982). In these studies the subjects' reaction time is determined for a varying number of choice alternatives, usually from one to eight. The rate of increase in reaction time with an increasing number of stimuli is seen as an index of intelligence; a low rate is associated with a high IQ. The larger increase in reaction time of black Americans in comparison with whites has been taken as evidence for genuine intergroup differences in intellectual functioning (Jensen, 1985). In an analysis of variance with cultural group and the number of choice alternatives as the independent variables and reaction time as the dependent variable, the intergroup difference in the rate of increase will be reflected in the stimulus by culture interaction.

The usual interpretation of the C- and SC-components is debatable also for psychometric reasons. Item bias effects are not restricted to the SC-component, but also can give rise to an intergroup difference in total score and hence to an increase in the main effect for culture (see also Van de Vijver & Poortinga, 1985). In such a case both the SC-component and the C-component reflect bias.

This is illustrated in Figure 1. The figure is based on a study with simulated data carried out by the second author. Forty item scores were generated for 100 subjects in each of two groups. Thereafter bias was introduced for the item scores in one of the groups, first for 1 item, then for 2 items, and so on, up to all 40 items. In other words, a systematic disadvantage was introduced for one of the groups that increased as bias was imposed on more items. The study was done two times, once with a small value for bias and once with a large bias value. More details on the procedure are given in the appendix.

The results of the study are presented in Figure 1. The size of the SC- and C-component is evaluated in terms of generalizability coefficients, indicated with the symbol $\hat{\rho}^2$ (cf. Van de Vijver & Poortinga, 1982). The interpretation of these coefficients is straightforward; the larger the coefficient, the larger the contribution of the particular component to the test score variance. In Figure 1 we see that both the C- and the SC-

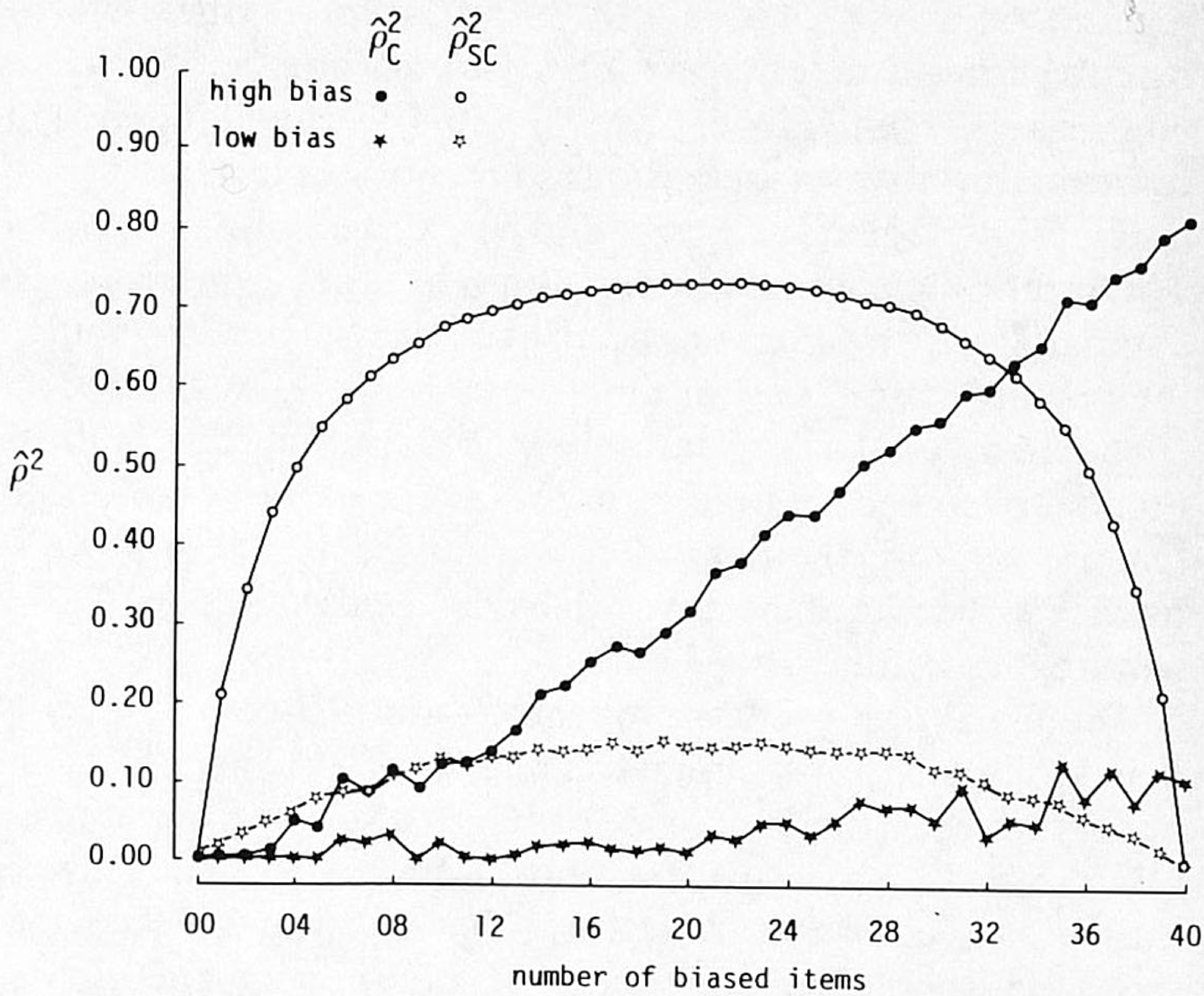

Figure 1: Estimated size ($\hat{\rho}^2$) of the SC-interaction and the main effect for C as a function of number of biased items for low bias and high bias.

component are about equal to zero when none of the items is biased. We can also see what happens with the two components when the number of biased items in a test gets larger; after an initial increase, the generalizability coefficient of the SC-component reaches an upper limit. After this point a further increase in the number of biased items does not lead any longer to a larger generalizability coefficient of the SC-component. Rather, with further increase in bias, the C-component becomes larger. One might say that the bias shifts to the C-component, that is, a cross-cultural difference in mean score, emerges. This shift is more pronounced in the condition with high bias, but is not restricted to it; for both low and high bias, the bias shift is observed. As an aside it may be noted that when the number of biased items is large, say three-quarters of the total, even iterative bias detection procedures (e.g., Mellenbergh, 1982) offer no solution for the problem of identifying the biased items in an instrument. In such a case the search for items that contribute substantially to the SC-component may lead to the removal of unbiased items.

It can be argued that the simulation study is unrealistic, because the bias systematically favored one group over the other. However, the situation in which one group is systematically advantaged by bias effects is more likely to occur than a situation in which each cultural group has an advantage on an approximately equal subset of the items. Cross-cultural studies often involve a comparison on tests constructed in a particular cultural, usually Western, context. In these comparisons it is hard to imagine that a group from a non-Western culture will be at an advantage on more than an occasional item.

As another criticism of this Monte Carlo study, it could be submitted that the analysis of variance model chosen here is a less than optimal choice and that the use of item response theory (cf. Lord, 1980) or χ^2 approaches (cf. Marascuilo & Slaughter, 1981) would be called for. However, these approaches will meet the same problems, as they all make implicitly or explicitly the assumption that bias affects the test only at item level. For example, in conditional item bias procedures (e.g., Mellenbergh, 1982) the total test score of a subject is used as a sufficient statistic for the estimation of the latent trait. When bias systematically favors one group over another, the total test score has a different meaning across groups. In view of the considerations mentioned, it seems to us that the removal of systematic bias by means of item bias methods has a more than casual resemblance with the famous baron of Münchhausen who allegedly pulled himself out of a swamp by his hair.

Returning to the ANOVA model, we have to point out that so far we have emphasized the SC- and the C-components. If bias is seen as the consequence of inappropriate sampling from a domain of stimuli or, for that matter, a population of subjects, similar arguments can be put forward in respect to other factors contributing systematic variance to the dependent score variable. For example, when in one of the cultural groups involved, the test taps the intended psychological construct in only a subgroup of the subjects, the effect for persons (nested in culture) will be affected. Rather than providing a list of examples in which the interpretation of bias is problematic, we shall draw the obvious conclusion that the detection of bias is more complicated than usually assumed in current approaches to item bias analysis and that each systematic component of Equation 1 can be affected by bias.

AN ALTERNATIVE MODEL?

An alternative model may be suggested as more appropriate. Omitting subscripts it can be written:

$$X = \mu + C^* + S^* + (P, PC)^* + (SC)^* + (SP, SPC)^*, E \qquad [2]$$

where: $C^* = C_t + C_b$

$S^* = S_t + S_b$

etc.

The subscript *t* stands for true; it indicates a genuine difference in the universe of generalization. The subscript *b* refers to bias, that is, a difference observed in the data set at hand, which is *not* to be found in the universe of generalization.

This model, more than the one presented in Equation 1, makes clear that the presence of bias in the dependent variable cannot be ruled out on the basis of only psychometric evidence. The reason is the impossibility of estimating all the effects in the extended model simultaneously. Thus the model that seems to be necessary from a theoretical point of view is impractical and does not allow for an estimation of components without imposing highly restrictive and theoretically often undesirable assumptions.

BEYOND BIAS ANALYSIS

It may seem to follow from the present position that no cross-cultural comparison is possible, as each component can represent either bias or genuine differences. However, all cross-cultural comparisons are not equally open to challenge and the plausibility of interpretations will vary widely.

It is primarily the responsibility of researchers to protect their inferences against plausible alternative explanations (e.g., Poortinga & Malpass, 1986). An obvious way to do this is by analyzing the effects of likely sources of bias. Such an analysis is possible only when relevant factors are included in the design of a study. Which of the various potential sources of bias form a serious threat to a meaningful interpretation of the results is to be determined by the researcher prior to data collection on the basis of existing knowledge and theoretical considerations. As a consequence, the explanation of cross-cultural differences becomes the focus of interest, whereas the distinction between bias and genuine differences becomes less prominent. In the analysis of bias the central issue is to distinguish valid cross-cultural differences from measurement artifacts, but no further attempt is made to identify the cultural antecedents in terms of which observed bias can be explained (Poortinga & Van der Flier, 1987). This is not to say that bias analysis is in any way superfluous. Whenever an instrument is used with a view to compare the results across cultures, it is essential that these results are equivalent. Bias analysis provides means to check on this. However, the a priori specification of a relevant explanatory variable is to be preferred over the a posteriori detection that some unknown variable lies behind an intercultural difference.

EXPLAINING CROSS-CULTURAL DIFFERENCES

In cross-cultural research interest is focused on the explanation of the variance accumulated in the factor culture. The observation of intergroup differences can only be the starting point, and the task of the cross-cultural psychologist is not finished before the intergroup variation in scores can be attributed to specific antecedent variables. In general, an investigator is more successful when a large part of the variance accumulated in the factor culture is explained by such variables. We shall indicate a variable that is introduced in a study for the purpose of explaining cross-cultural differences with the term *context variable*. In this section we shall describe a methodological-statistical framework in which the adequacy of context variables can be evaluated. For reasons of simplicity the case of only one context variable will be considered, but generalizations to a larger number are straightforward.

It is essential that prior to the data collection an investigator hypothesizes what will be the relevant context variables to explain the observed differences on a particular target variable and develops appropriate measurement procedures for these variables. In other words, context variables have to be explicitly introduced in the design of a study.

Our treatment of a data set containing information about a context variable and a target variable consists of two steps. In the first step, the contribution of the context variable to the variance in the target variable is calculated, whereas in the second step it is investigated whether there are any cross-cultural score differences remaining after the contribution of the context variable to the variation in the target variable has been eliminated.

For the data analysis a hierarchical regression model is used (good descriptions of the model can be found in Pedhazur, 1982, and in Cohen & Cohen, 1983). Statistically the approach can be presented in the form of two regression equations, in which the target variable is the dependent variable and both the factor culture and the context variable are treated as independent variables. The first is the regression equation expressing the contribution of the context variable to the dependent variable. It can be expressed as follows:

$$X_{pk} = a + b_k K_p + E_{pk} \qquad [3]$$

in which:

X_{pk} is the score on dependent variable X of individual p with value k on the context variable K;
a is the intercept;
b_k is the regression coefficient of the context variable K;
K_p is the score of individual p on the context variable K;
E_{pk} is the error variable.

It is important to note that the factor culture does not enter this equation. In this analysis the explanatory power of the context variable for the scores on the dependent variable is investigated, without taking culture as a separate factor into account. The presumed appropriateness of the context variable can be evaluated by means of the squared multiple correlation coefficient denoted by R^2, which indicates the proportion of variance in the target variable accounted for by the context variable. This coefficient can be tested for significance (e.g., Cohen & Cohen, 1983, formula 3.6.1 or Pedhazur, 1982, formula 3.21). When R^2 differs from zero, the context variable can be taken to contribute to the explanation of score variation in the dependent variable. The higher R^2, the more powerful a context variable. On the other hand, when R^2 does not differ from zero, the context variable and the dependent variable can be taken to be independent.

When the multiple correlation of this analysis differs from zero, the next stage of the analysis can be carried out. At this stage it is investigated whether any cross-cultural difference in the dependent variable remains after the effect of the context variable has been taken into account. For this purpose culture is entered as an independent variable. The model equation is given by:

$$X_{p(c)k} = a' + b_k K_p + b_{c,k} C_c + E_{p(c)k} \qquad [4]$$

where:

> $X_{p(c)}$ is the score on the dependent variable with person p nested in culture c;
> a′ is the intercept;
> $b_k\ K_p$ is defined as in [3];
> $b_{c,k}\ C_c$ is the regression effect for culture, corrected for the effect of the context variable;
> $E_{p(c)k}$ is the error term.

Some additional comment on the terms in this equation and the implications of this analysis may be useful. Culture is a nominal variable; this impedes its direct introduction in the regression analysis. Techniques have been developed to deal with nominal variables in regression analyses, such as dummy coding or effect coding (details can be found in Cohen & Cohen, 1983, chap. 5, or Pedhazur, 1982, chap. 9).[1]

The crucial term in the equation is $b_{c,k}\ C_c$, that is, the effect of culture after the effect of the context variable has been accounted for. When part of the intergroup differences is unaccounted for by a context variable, the variable culture will still contribute to the variation in the dependent variable. In this case a significant increment in R^2 will be observed in the second analysis (in comparison with the first). The increment can be tested for significance (e.g. Cohen & Cohen, 1983, formula 4.4.1 and 4.4.2). In other words, a significant increase in R^2 from the first to the second analysis indicates that there still are important sources of cross-cultural differences not yet explained in terms of the context variable.

On the other hand, when the increase in R^2 is not significant (or so small as to be of no practical meaning), the cross-cultural differences have been adequately dealt with. In that case the cross-cultural score differences have been fully explained by the context variable.

Not every context variable that gives rise to a significant R^2 in the first analysis will explain part of the variance on the factor culture. One can imagine a context variable that correlates highly with the dependent variable but leaves intergroup differences unaffected, that is, a context variable that explains interindividual variance but not intercultural variance. Apart from the question of how likely it is that such a variable will be found in practice, it should be noted that in this way a context variable is identified that cannot be the antecedent of intergroup differences on the dependent variable.

In the presentation of the framework, the impression may have been created that an investigation is more successful as more variation in the dependent variable is explained. This need not be the case. The discovery that a context variable is unrelated to the dependent variable and to any group difference can be very valuable. To understand this argument one may think of the concepts of convergent and discriminant validity (Campbell & Fiske, 1959), two aspects of validity that contain complementary information.

In summary, regression analysis (or a similar multivariate technique to reduce the variance attributable to the factor culture) is an important tool in the testing of presumed antecedents of intergroup differences, as it stimulates the researcher to formulate precise hypotheses.

AN IMPORTANT COROLLARY

The procedure proposed here has some interesting implications. If a context variable "explains" part of the variance due to the factor culture, the effect for culture computed in the analysis will become smaller. If another context variable explains part of the remaining variance, the effect for culture again will be reduced. Thus the approach described here can be seen as a psychometric translation of the well-known suggestion by Whiting (1976) that cross-cultural research amounts to the unwrapping of the "packaged variable" culture. In the ideal study the set of context variables will be chosen in such a way that the remaining effect for culture will be zero.

An important corollary of the procedure is that the effect for culture will come closer to zero as the explanation of cross-cultural differences is more successful. It may seem somewhat paradoxical, but the consequence of our argument is that a cross-cultural psychologist is not interested in the variable culture per se, but only in specific context variables that can explain observed differences on some dependent variable.

An additional point to be noted is that within a particular study, the interpretation dilemma mentioned in the introduction is solved when the variance in the factor culture has been reduced to zero. On the other hand, as long as not all variance can be explained the dilemma persists.

A DIGRESSION ON CONTEXT VARIABLES

Cross-cultural studies frequently are based on explicit measurement of only the dependent variable. Much less attention is paid to the independent variable. This variable is usually specified a priori, but not explicitly measured. For example, a researcher interested in the relationship between the quality of school education and Piagetian conservation will select cultures for data selection on the basis of their presumed differences on the variable "schooling." In this kind of research, additional and unintended cross-cultural differences beyond the difference in school quality are often not recognized; observed intergroup differences are exclusively ascribed to differences in school quality. Other variables on which the cultures differ include, for example, nutritional status, parental style, and degree of Westernization. The shortsightedness of only considering one presumably relevant dimension at which the cultures differ has been criticized by Campbell (e.g. Campbell & Naroll, 1972). He pointed out that alternative explanations often cannot be ruled out and should be taken into account.

One of the ways in which this can be achieved is by introducing a more detailed analysis of the factor "culture." It will be possible to rule out alternative hypotheses only when these have been investigated. The inclusion of context variables in a study and the investigation of their influence on some target task can be considered a methodological elaboration of Campbell's argument.

Intergroup score differences can be the result of a large variety of antecedent factors. In the present approach no constraints are imposed on the kind of variables that can form meaningful context variables. They can belong to the psychological domain, but economic variables, such as the per capita income, or sociological variables, such as socioeconomic status, are equally pertinent.

In some instances it will be possible to determine for each subject an individual score on the context variable, whereas in other cases all subjects from a culture will be given an identical score. For example, in a study on the relationship between cognitive style and socialization practices, the investigator can gather data about these practices (the context variable) in two ways, either by collecting information from existing sources such as the Human Research Area Files or by gathering data on each subject in the actual sample. In the first case each member of a culture will be assigned the same score and there will be no intracultural variation on the contextual variable, whereas in the second case interindividual differences are very likely to occur. In the framework we are suggesting both approaches are permitted.

Nevertheless, there are some constraints on context variables. First, variables measured on a nominal scale cannot be used as context variables. We can illustrate this with an example. Suppose that two cultural groups differ in average score on some instrument. If this is attributed to the fact that they speak a different language, this variable can be entered in the first regression equation, as a dummy variable. If in the second equation the levels in the factor culture are given the same values as the corresponding languages, all variance in the factor culture is explained. The same holds for race, climate, and any other variable that forms a nominal scale. This is due to the fact that any transformation preserving the identity of the score categories is admissible. The scores can always be transformed to coincide with the set of scores on another nominally scaled variable, including the factor culture itself. Therefore, the variance of all imaginable context variables is confounded in comparisons in which context variables are measured only at the nominal level. These variables can be exchanged for each other freely. When a context variable is measured at a higher level, this problem does not arise. It could even be argued that the present approach amounts to an attempt at replacing the nominal variable culture by a number of its constituent elements that have a higher measurement level.

The second constraint has to do with the explanation of the variance that a context variable and the dependent variable have in common. Psychometrically, the most effective way to deal with the variance in the factor culture is to administer two parallel instruments, of which one is considered as the dependent variable and the other as a context variable. It is obvious that in a psychological sense, nothing has been explained in this case. A similar objection applies when a dependent variable and a context variable can be related to each other through a superordinate variable of which both are merely specific instances. For example, the difference between two groups in reading ability cannot be satisfactorily explained with reference to a corresponding difference in arithmetical skills, if both can be seen as a function of intergroup differences in the number of years of formal education.

In the kind of analysis that is the subject of discussion, a context variable can also serve to demonstrate that observed cross-cultural differences cannot be attributed to method variables. If a context variable and a dependent variable are based on a common method but differ in some crucial aspect, the context variable can be used to control for the event that an observed cross-cultural difference is due to method variance. In the next section we shall provide an example of this possibility.

The two-fold usage of context variables suggested here presupposes at least a modicum of theoretical insight during the design phase of a study and a careful analysis of the interrelationships of various context variables after the data have been collected.

AN EXAMPLE

The use of context variables will be illustrated with an example from a cross-cultural study on the cultural invariance of basic personality parameters. The experiment concerned was on the habituation of the orienting reflex or OR (Poortinga, 1986). The skin conductance response (SCR) was the dependent variable. Pure tones of 500 Hz and 1-sec duration were presented to subjects at 20-sec intervals. There were two identical sessions for each subject. Results for four samples (two groups from each of two cultures: Indian students, illiterate Indian tribals of the Juang group, Dutch students, and Dutch military conscripts) are presented in Figure 2. The SCR scores were highest for the Indian students and some significant intercultural differences were found for the level of the initial OR, although not for the main variable of interest, that is, the rate of habituation.

It is instructive for our argument to consider what would be concluded when these data would have been treated in the classical way, for example, by subjecting them to an ANOVA. Such an analysis was carried out, separately, for each of the two sessions. The results are presented in Table 1. In the first session no significant effects were found; in the second session both culture and group turned out to be significant. Various admittedly speculative, but not implausible, explanations of the observed differences in SCR levels could be generated. For example, according to Nebylitsyn (1972, p. 70), a high initial OR is indicative of a high dynamism in the excitation process triggered in the nervous system by stimulation.

However, the psychological meaning of the observed difference in initial OR was far from clear. The question could be asked whether the intercultural differences in SCR should be taken as an index of some underlying process or as due to

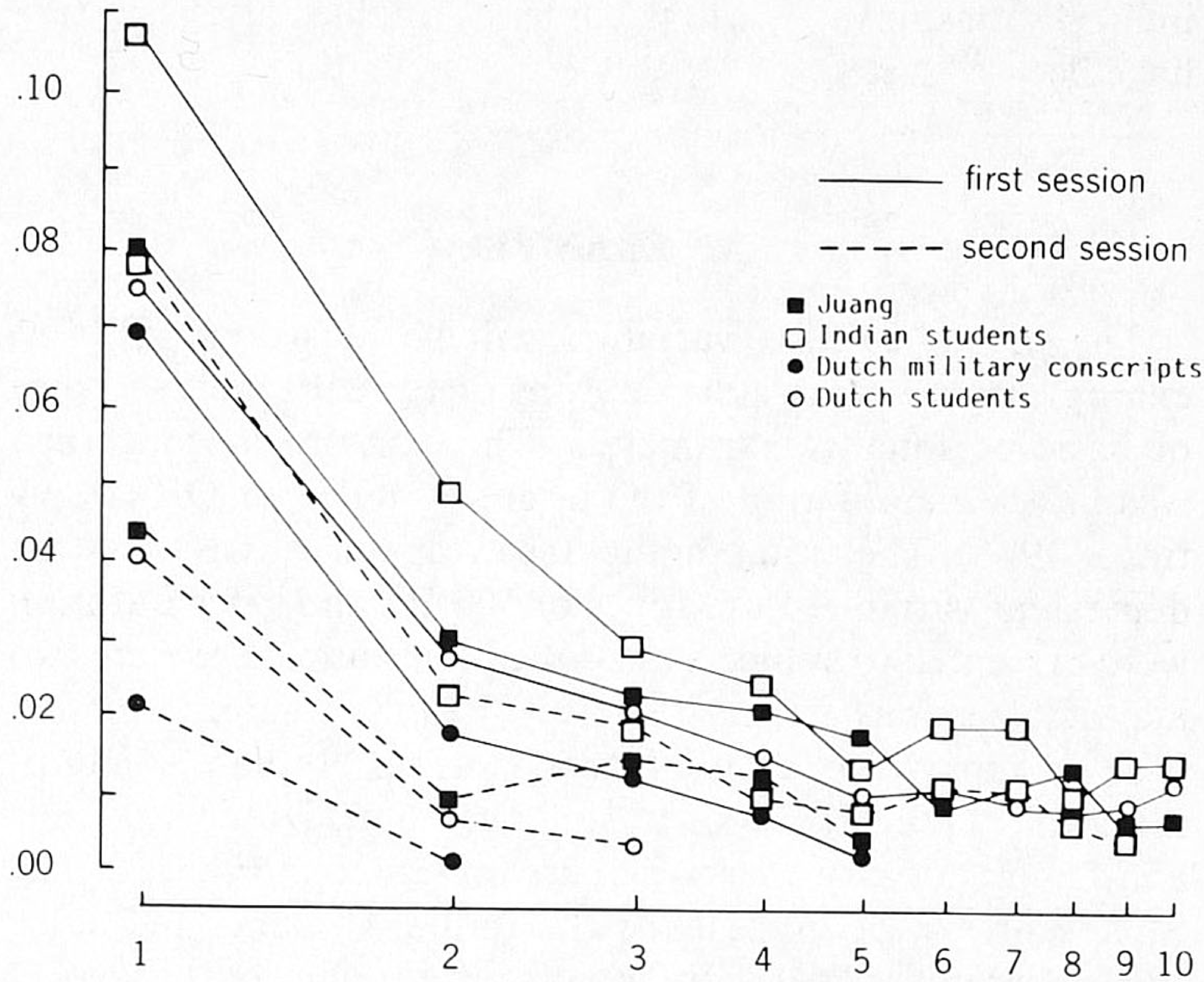

Figure 2: OR-habituation expressed in skin conductance response (log SCR μmho + 1.0), averaged over subjects, as a function of trial number. A graph is discontinued if < 20% of the subjects showed a noticeable response on a trial (after Poortinga, 1986).

method artifacts. In the project, various measurements had been collected on the "state" of the subjects in the experimental situation. The most obvious choice for a context variable, given an SCR measure as the dependent variable, was the extent of spontaneous fluctuations in skin conductance recorded during periods of rest at the beginning and at the end of each experimental session.

First, the question needs to be asked whether this variable meets the constraints mentioned in the previous section. The level of measurement does not lead to any problem, because the scores were a logarithmic transformation of a physical variable, namely, changes in the electrical conductance of the subject's skin. The scores for the OR trials and the rest

TABLE 1
Significance Levels for Effects of Culture, Group, and Culture by Group Interaction in an Analysis of Variance

Source	Session 1	Session 2
Culture (C)	.19	.01*
Group (G)	.71	.01*
C x G	.10	.42

*$p < .05$.

condition have many components of method variance in common, but differ in one essential aspect, namely, the presence of the stimulus tone versus the absence of an external stimulus. Therefore, the scores during rest could be used to analyze whether the difference in the OR was indeed due to the differential impact of the stimulus tone on people belonging to different cultures.

In the first of the two regression analyses (cf. Equation 3), the impact of the rest SCR on the dependent variable (i.e., the OR) was investigated. For each of the two sessions, a separate analysis was carried out; the results are presented in Table 2. The multiple correlation coefficient for rest SCR was found to be highly significant ($p = .00$) in both sessions. In the second regression analysis (cf. Equation 4), an additional set of independent variables was introduced: culture, group, and the culture by group interaction (analogous to the above mentioned ANOVA). By means of this hierarchical regression analysis it was shown that no intercultural difference remained for the first OR trial after the differences in rest SCR had been accounted for. The increase in the squared multiple correlation coefficient from the first to the second analysis was less than 0.03 in each of the two sessions. Therefore, the variance in the factor culture, still important in the ANOVA, was reduced to nonsignificance, thereby preempting any "cultural" interpretation of the differences in the orienting reflex.

As an aside it may be noted that in view of other results in the same study, the differences in skin conductance responses were

TABLE 2
Squared Multiple Correlation Coefficients of Various Effects in a Regression Analysis Procedure with Rest-SCR Scores as a Context Variable and the Initial Orienting Reflex as the Dependent Variable

	Session 1	Session 2
Squared Multiple Correlation (rest)	.382*	.210*
Increments in Squared Multiple Correlation (second analysis)	G .013 GxC .014 C .015	G .007 C .021 GxC .022

*$p < .01$.

attributed to intercultural differences in the "arousal" or "anxiety" evoked by the experimental situation.

CONCLUSION

In an earlier article (Van de Vijver & Poortinga, 1982), we emphasized that bias is not necessarily restricted to the stimulus by culture interaction in an ANOVA design. The argument has been extended here, indicating how any systematic component in a design can be affected by bias. If some restrictive assumptions hold, the presence of bias can often be demonstrated quite unambiguously. However, the demonstration that bias effects have *not* played a role is far more difficult. It can remain quite unclear whether an observed intercultural difference is valid, or due to bias, even if the data do pass one of the usual tests for bias.

Researchers can improve the interpretation of results by incorporating additional variables, called context variables, in the design of a study. If a relevant variable is introduced, the variance attributable to the psychologically unspecified factor culture will be reduced. A complete explanation of the intercultural differences has been given, if the remaining variance in the factor culture is reduced to zero. The researcher has to be aware that meaningful context variables have to meet certain conditions, for example, that they should not form nominal scales.

It is obvious that bias analysis and context variable analysis in many respects are complementary. If an SC interaction virtually disappears with the elimination of only a few items, the interpretation of intergroup differences may well improve. On the other hand, if a substantial proportion of the items appears to be biased, their elimination may obscure valid differences. Therefore, it would be erroneous to suggest that context variable analysis should always be preceded by bias analysis. For context variable analysis the distinction between bias and valid differences is not of primary importance.

In this article culture is considered as an independent variable, in a sense similar to that suggested by Segall (1983, 1984). The successful elimination of culture as a source of variance means that the concept of culture—which post hoc can be used to "explain" any observed intergroup difference—has been superseded by variables with a more focused meaning. In this way a more precise interpretation of so-called cross-cultural differences can be obtained.

APPENDIX

The Monte Carlo study was meant to simulate a data set conforming to the common cross-cultural research design in which a number of stimuli are administered to two groups of subjects (cf. Equation 1). A data matrix was generated representing a set of 40 stimuli administered in two different cultural groups of 100 subjects each. The generating process consisted of a number of steps. First, a vector of 40 stimulus parameters and two vectors of 100 person parameters, one for each "cultural group," were drawn from a standard normal distribution.

On the basis of these vectors the score for each subject on each stimulus—the cell entries of the data matrix—were computed using the Rasch model (Rasch, 1960; Lord, 1980). To each score a random error component was added, drawn from a normal distribution with a mean of zero and a standard deviation of 0.25. Finally, to make an item biased, a constant was added to all the entries of the corresponding column in the data matrix of one cultural group.

Data were generated for two values of the bias component, 0.1 (low) and 0.7 (high). The number of biased items was systematically varied, from 0 to 40, constituting 41 different bias conditions. These were studied, both under low and high bias. For each of these 82 conditions, 25 computer runs were carried out.

It may be noted that the data were not dichotomized at any stage; the data matrix was not meant to simulate a test with dichotomous items as the use of the Rasch model might suggest. Consequently, cell entries could take values other than 0 and 1.

NOTE

1. For example, in effect coding the n_c cultures involved in a study are represented by $n_c - 1$ independent variables. Each of these contains three values, −1, 0, and + 1. The first independent variable will contain a value of 1 for all individuals of an arbitrarily chosen first culture, 0 for the individuals of all other groups, except an arbitrarily chosen last one, which is coded as −1. The vector of the second independent variable will have a 1 for the individuals of the second culture, a 0 for all other cultures, again with the exception of the final culture that is coded as −1. Although effect coding is only one of the possible ways of introducing nominal variables, it may be noted that the size of R^2 is not affected by a particular choice of coding.

REFERENCES

Berk, R. A. (Ed.). (1982). *Handbook of methods for detecting item bias.* Baltimore: Johns Hopkins University Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Campbell, D. T., & Naroll, R. (1972). The mutual methodological relevance of anthropology and psychology. In F. L. K. Hsu (Ed.), *Psychological anthropology* (rev. ed., pp. 435-463). Cambridge MA: Schenkman.

Cleary, T. A., & Hilton, T. L. (1968). An investigation of item bias. *Educational and Psychological Measurement, 28,* 61-75.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Jensen, A. R. (1982). Reaction time and psychometric g. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 93-132). Berlin: Springer Verlag.

Jensen, A. R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *The Behavioral and Brain Sciences, 8,* 193-263.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum.

Marascuilo, L. A., & Slaughter, R. E. (1981). Statistical procedures for identifying possible sources of item bias based on χ^2-statistics. *Journal of Educational Measurements, 18,* 229-248.

Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics, 7,* 105-118.

Osgood, C. E. (1977). Objective cross-national indicators of subjective culture. In Y. H. Poortinga (Ed.), *Basic problems in cross-cultural psychology* (pp. 200-235). Lisse: Swets and Zeitlinger.

Nebylitsyn, V. D. (1972). *Fundamental properties of the human nervous system.* New York: Plenum Press.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart & Winston.

Poortinga, Y. H. (1971). Cross-cultural comparison of maximum performance tests. *Psychologia Africana Monograph Supplement, 6.*

Poortinga, Y. H. (1986). *Psychic unity versus cultural variation: An exploratory study of some basic personality variables in India and The Netherlands.* Unpublished report, Tilburg, Tilburg University.

Poortinga, Y. H., & Malpass, R. S. (1986). Making inferences from cross-cultural data. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 47-83). Newbury Park, CA: Sage.

Poortinga, Y.H., & Van der Flier, H. (1986). The meaning of item bias in ability tests. In S. H. Irvine & J. W. Berry (Eds.), *The cultural context of human abilities.* Cambridge: Cambridge University Press.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Nielson and Lydicke.

Segall, M. H. (1983). On the search for the independent variable in cross-cultural psychology. In S. H. Irvine & J. W. Berry (Eds.), *Human assessment and cultural factors* (pp. 127-138). New York: Plenum.

Segall, M. H. (1984). More than we need to know about culture, but are afraid not to ask. *Journal of Cross-Cultural Psychology, 15*, 153-162.

Van der Flier, H. (1982). Deviant response patterns and comparability of test scores. *Journal of Cross-Cultural Psychology, 13*, 267-298.

Van der Flier, H., Mellenbergh, G. J., Adèr, H. J., & Wijn, M. (1984). An iterative item bias detection method. *Journal of Educational Measurement, 21*, 131-145.

Van der Vijver, F.J.R., & Poortinga, Y. H. (1982). Cross-cultural generalizability and universality. *Journal of Cross-Cultural Psychology, 13*, 387-408.

Van de Vijver, F.J.R., & Poortinga, Y. H. (1985). A comment on McCauley and Colberg's conception of cross-cultural transportability of tests. *Journal of Educational Measurement, 22*, 157-161.

Whiting, B. B. (1976). The problem of the packaged variable. In K. F. Riegel & J. A. Meacham (Eds.), *The developing individual in a changing world* (Vol. 1, pp. 303-309). Chicago: Aldine.

Ype H. Poortinga is on the staff of Tilburg University (Department of Psychology, Tilburg University, Tilburg 5000 LE, The Netherlands). He is currently Vice President and President-Elect of the International Association for Cross-Cultural Psychology.

Fons J. R. van de Vijver has a research position at Tilburg University, The Netherlands.